**Introduction**

The objective of this study of protein-protein interactions is to introduce a generic computational model to exploit the synergy among various protein information. The computational model is based on meta learning. It has the flexibility to integrate various component prediction tools into a unique prediction system.

For the purpose of demonstration, we provide the executable code of PPI-MetaGO and the data used in this study. The user can prepare his(her) own training and test datasets for PPI prediction.

**Availability**

The complete code and data are available on the Google Drive through this [link](https://drive.google.com/drive/folders/127bjTMPY68Q_VWpwhBPQoCYNOJ11JqIG?usp=sharing). For the purpose of a quick test drive of PPI-MetaGO, we have prepared the sample training and test data constructed from three protein pair datasets, EC2, AT and SC5, available at <https://github.com/mlbioinfolab/ppi-metago>. In addition, the executable code of PPI-MetaGO is also available, and the reader can test-drive PPI-MetaGO directly on these preprocessed datasets.

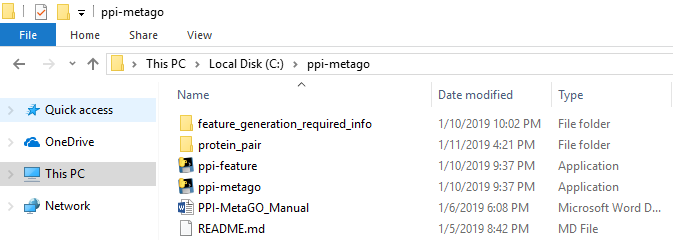
**Materials**

***Computer code:***

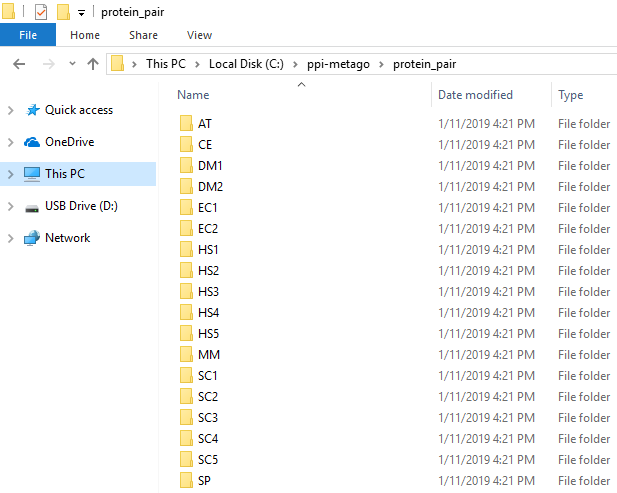
1. ppi-metago.exe: executable code of PPI-MetaGO
2. ppi-feature.exe: executable code of the feature generator that generates the feature values of protein pairs for training and testing.

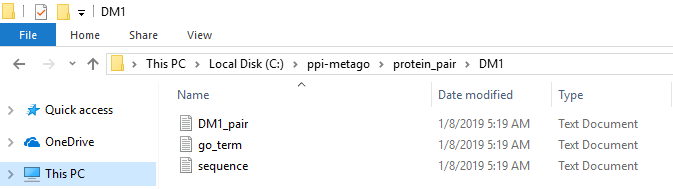
***Data folders/subfolders:***

1. main\_folder: a folder that stores ppi-metago.exe, ppi-feature.exe, protein pair data, protein sequences, and GO terms. After unzipping the zipped file downloaded from the Google Drive, the contents of a main folder, e.g. ppi-metago, will look like what the snapshot below shows.



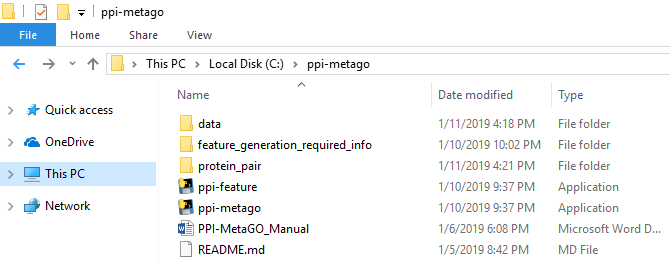
1. protein\_pair: a folder that stores the information of protein pairs from various species, each organized in a subfolder. The two snapshots below show 18 subfolders, and each subfolder stores GO terms, protein sequences, and protein pairs.





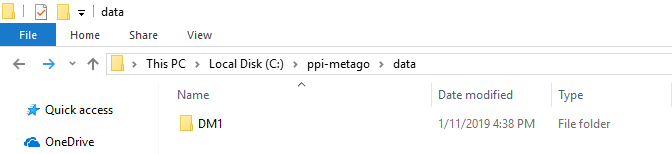
1. feature\_generation\_required\_info: a folder that stores the information required of ppi-feature.exe, including all the protein sequences used in this study and the complete collection of GO terms, to generation the protein pair features.

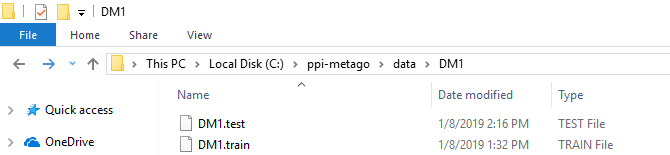
1. data: a folder that stores the training data and the test data in their feature vector form, produced by ppi-feature.exe for training PPI-MetaGO and predicting PPIs.



**NOTE.** The folder **data** is automatically constructed after running ppi-feature.exe to generate protein pair features.

Take DM1 for example.



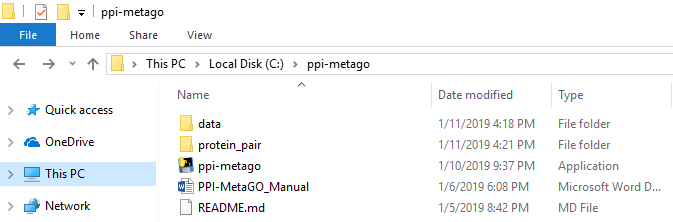


The details of running ppi-feature.exe are provided in the next section.

**Instructions to use the meta classifiers to predict the protein-protein interactions**

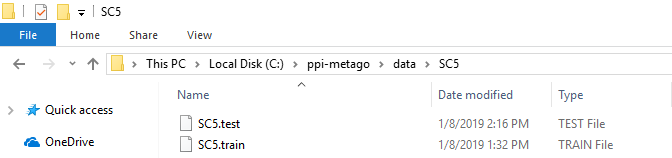
*A. For a quick test drive of PPI-MetaGO, using code and data on GitHub:*

1. Create a main folder, e.g. **ppi-metago**, and subfolders, e.g. **protein\_pair** and **data**, on your PC (Windows system). From GitHub, download ppi-metago.exe and place it in **ppi-metago**. Download the datasets in one of the subfolders in the “data” folder from GitHub, e.g. SC5.train and SC5.test, and place them in **data** on your PC.



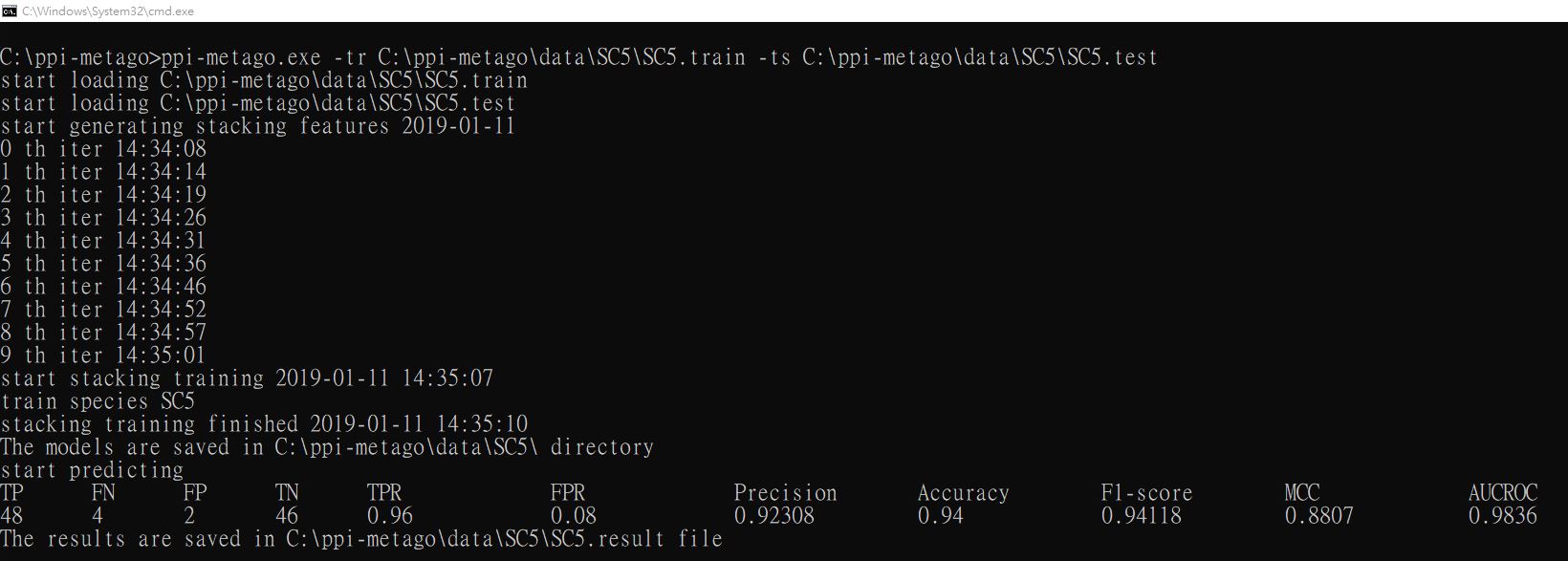
1. How to run PPI-MetaGO to train and test PPI prediction models in Windows Systems?
   1. To complete both training and testing in one step by running ppi-metago.exe, type the following command. Refer to the snapshot below.

Take SC5 for example.

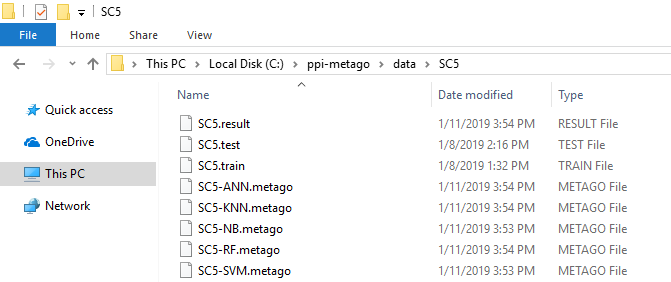


Type the following.

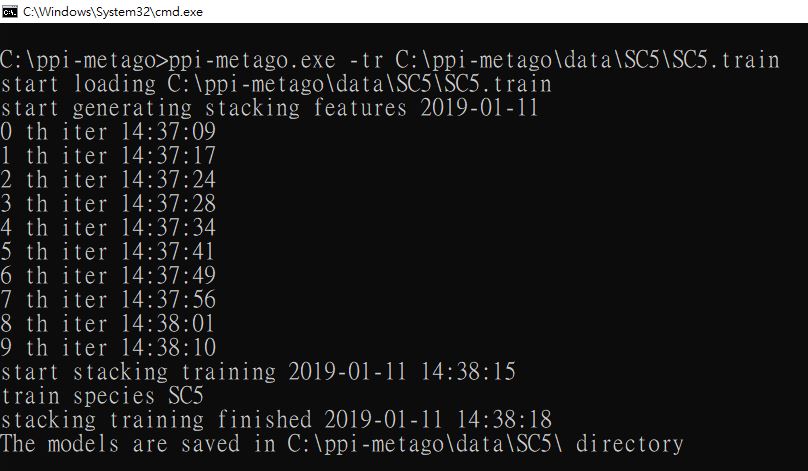
ppi-metago.exe –tr c:\ppi-metago\data\SC5\SC5.train –ts c:\ppi-metago\data\SC5\SC5.test



The prediction performances, e.g. Accuracy, MCC, AUCROC, are output on the screen. In addition, one result file and five trained models are produced and stored in the SC5 subfolder.

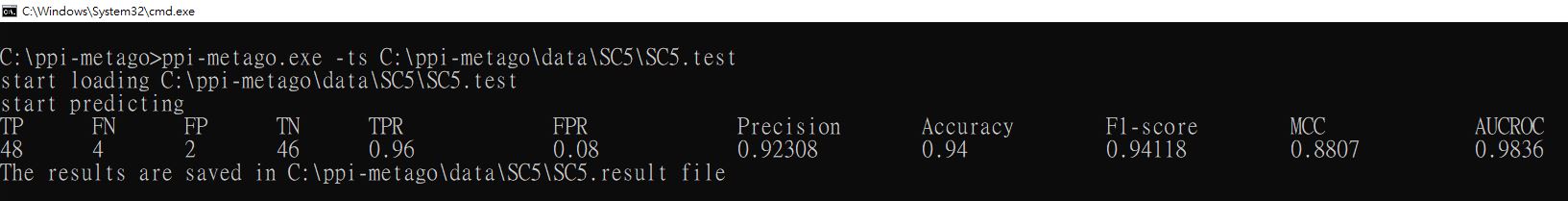


* 1. To train prediction models only w/o running tests by running ppi-metago.exe, type the following command.



In this case, no test result is produced while five prediction models will be trained and stored to the SC5 subfolder.

* 1. You can also used the previously trained models to run tests, e.g. type the following command.

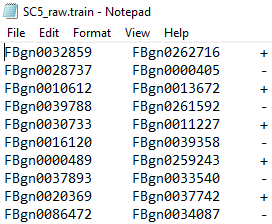
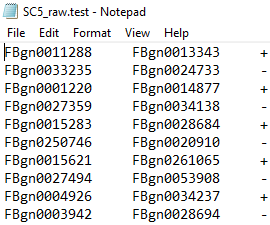


**NOTE.** For a quick test drive of PPI-MetaGO, there is no need for running ppi-feature.exe to generate training and test data in their feature vector forms. We have already prepared the vectorial training and test data and placed them in the “data” folder on GitHub.

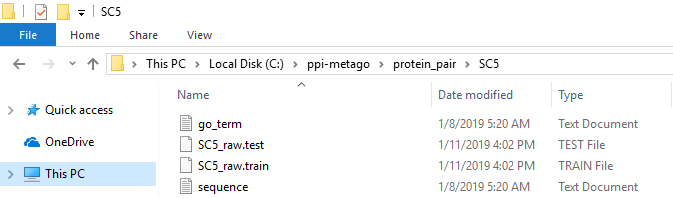
*B. How to generate protein pair features and prepare training/test data in the feature vector form?*

1. Prepare the training and test datasets of protein pairs in the following form.

Take SC5 for example.

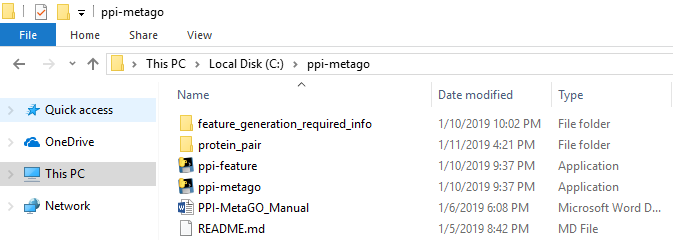
 

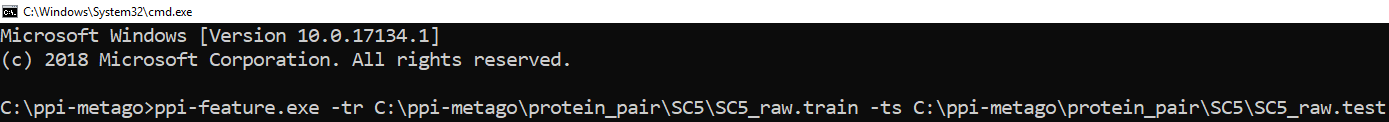
Store them in the SC5 subfolder in the protein\_pair folder. See the snapshot below.



1. Run ppi-feature.exe on SC5\_raw.train and SC5\_raw.test to generate the vectorial training and test datasets. See the following snapshot.

**Note.** You need to place the feature\_generation\_required\_info folder with ppi-feature.exe in the main folder. See below.





After running ppi-feature.exe, we can obtain the vectorial training and test files, as shown below, and stored in the data folder.

